Effect of Lossy Vector Quantization Hyperspectral Data Compression on Retrieval of Red-Edge Indices

Shen-En Qian, Allan B. Hollinger, Melanie Dutkiewicz, Herbert Tsang, Harold Zwick, and James R. Freemantle

***Abstract*—This paper evaluates lossy vector quantization-based hyperspectral data compression algorithms, using red-edge indices as end-products. Three compact airborne spectrographic imager (CASI) data sets and one airborne visible/infrared imaging spectrometer (AVIRIS) data set from vegetated areas were tested. A basic compression system for hyperspectral data called the "reference" system, and three-speed improved compression systems called systems 1, 2, and 3, respectively, were examined. Five red-edge products representing the near infrared (NIR) reflectance shoulder (Vog 1), the NIR reflectance maximum (Red_rs), the difference between the reflectance maximum and the minimum (Red_rd), the wavelength of the reflectance maximum (Red_lo), and the wavelength of the point of inflection of the NIR vegetation reflectance curve (Red_lp) were retrieved from each original data set and from their decompressed data sets. The experiments show that the reference system induces the smallest product errors of the four compression systems. System 1 and 2 perform fairly closely to the reference system. They are the recommended compression systems since they compress a data set hundreds of times faster than the reference system. System 3 performs similarly to the reference system at high compression ratios. Product errors increase with the increase of compression ratio. The overall product errors are dominated by Vog 1, Red_rs, and Red_rd, since the amplitude of product error for these products is over one order of magnitude greater than those for the Red_lo and Red_lp products. The difference between the overall error from the reference and that from system 1 or 2 is below 0.5% at all compression ratios. The overall product error induced by system 1 or 2 is below 3.0% and 2.0% for CASI and AVIRIS data sets, respectively, when the compression ratio is 100 and below. Spatial patterns of the product errors were examined for the AVIRIS data set. For all products, the errors are uniformly distributed in vegetated areas. Errors are relatively high in nonvegetated and mixed-pixel areas.**

***Index Terms*—Data evaluation, hyperspectral data, lossy data compression, red-edge, vector quantization.**

I. Introduction

A HYPERSPECTRAL imager (also called imaging spectrometer) acquires images in hundreds of narrow contiguous spectral bands, and produces a three-dimensional (3-D) datacube for a scene on the ground. Its raw data rate can easily exceed the available satellite downlink capacity or exhaust onboard storage capacity. Data compression is one solution to significantly ease the mission requirements for such a sensor. Vector Quantization (VQ) [1] has been proposed as a suitable method for lossy data compression, because of its promise of a high compression ratio (100 : 1 or more) and its relative simplicity [2], [3]. A significant drawback to this technique, in terms of operational use, is that it requires large computational resources, particularly for the codebook generation step. Since the size of a hyperspectral datacube can be hundreds of times larger than that of a traditional remote sensing image, the processing time required to train a codebook or to match a codevector can also be hundreds of times larger.

Successful strategies have recently been developed to create a fast hyperspectral VQ compression algorithm. For example, using the algorithm, spectral-feature-based binary coding (SFBBC) [4], both the codebook generation time and coding time are decreased by a factor of 30 to 40 at a loss of peak SNR (PSNR) of less than 1.5 dB. Another hyperspectral VQ compression algorithm made use of training set subsampling [5] to further reduce the codebook generation time. When the training set was formed by subsampling the datacube, for example, at a rate of 2%, the codebook generation time was reduced by a factor of around 50 at a loss of PSNR of less than 0.75 dB. A more efficient hyperspectral VQ compression algorithm was presented in [6], where a spectral index incorporated remote sensing knowledge into the compression system to improve the performance. A spectral index based multiple subcodebook algorithm (MSCA) was developed. For this algorithm, the scene of a hyperspectral datacube is delimited into n classes by segmenting its spectral index image. The datacube is then separated into n subsets corresponding to the n classes. The spectra in each class have similar spectral characteristics, and each class can thus be compressed with a small codebook. For the same total number of codevectors (i.e., the same compression ratio), n smaller and more efficient subcodebooks are generated instead of a single codebook. Each subcodebook is used to compress the spectra in the corresponding class. With the MSCA, both the codebook generation time and coding time can be improved by a factor of around n with almost no PSNR penalty. In [7], three hyperspectral VQ data compression systems that integrate the developed techniques were simulated and tested. The simulation results show that the codebook generation time could be reduced by more than three orders of magnitude, while the quality of the codebooks remained good. The overall processing speed of the compression systems could be improved by a factor of around 1000 at an average PSNR penalty of 1.5 dB.

In lossy data compression, the mean square error (MSE) is often used as a means to measure distortion of the reconstruction data and to compare coding techniques. In the compression of hyperspectral data, it is inadequate to evaluate the quality of

Manuscript received September 20, 2000; revised January 8, 2001.

S.-E. Qian and A. B. Hollinger are with the Canadian Space Agency, St. Hubert, PQ, J3Y 8Y9 Canada (e-mail: shen-en.qian@space.gc.ca).

M. Dutkiewicz, H. Tsang, and H. Zwick are with MacDonald Dettwiler and Associates, Richmond, BC, V6V 2J3 Canada.

J. R. Freemantle is with the Centre for Earth and Space Technology, Toronto, ON, M3J 3K1 Canada.

Publisher Item Identifier S 0196-2892(01)05502-4.

the reconstruction data using the MSE, since the end-users of hyperspectral data are likely to be image processing systems for specific remote sensing applications instead of human viewers. An appropriate distortion measure, which reflects the loss in scientific value of the original data in the context of the remote sensing applications of the data, is required. Unfortunately, such a quantitative measure does not exist either for specific applications or for general applications of hyperspectral data.

In order to evaluate data compression techniques used in remote sensing applications, studies of the impact of lossy data compression on remote sensing applications have been conducted and reported.

Qian *et al.* investigated the effects of their VQ hyperspectral data compression approach on the derivation of the fluorescence line height (FLH) in water applications [2]. In the evaluation, an FLH image was computed from the original datacube, and from the decompressed datacube at a compression ratio of 96 : 1. It was demonstrated that the compression system has a smooth effect, which eliminates some high frequency artifacts (salt and pepper noise). Three different classification regions were distinguished more clearly in the FLH image obtained from the decompressed data than in the FLH image obtained from the original data. In order to do a quantitative comparison, a homogenous area 25 $\times$ 25 pixels in size was extracted from each class from both the FLH images. The mean and standard deviation of each class were computed and analyzed for each area. It was shown that the means of the areas of the two FLH images were very close, but their standard deviations were very different. The standard deviations of the areas of the FLH image computed from the decompressed data were much smaller (around 1.0) than those from the original data (around 3.8).

Ryan and Arnold proposed a remote sensing application relevant distortion measure called the percentage maximum absolute distortion (PMAD) and evaluated their algorithm for land cover classification using the PMAD [8]. The classification accuracy of both the total image and its constituent classes varied in a regular way as the level of PMAD distortion increased. For the selected test datacube with a compression ratio of 17 : 1 at 4% PMAD distortion, the loss in classification accuracy for most classes was less than 8%. The total accuracy of classification was reduced from 96.8% for the original datacube to 82.8% for the compressed datacube.

Recently, Hu *et al.* investigated the effects of Qian *et al.*'s VQ hyperspectral data compression algorithm [4] on the retrieval of surface reflectance [9]. A compact airborne spectrographic imager (CASI) datacube of the boreal ecosystem-atmosphere study (BOREAS) [10] compressed at four compression ratios of 104 : 1, 75 : 1, 59 : 1, and 42 : 1 was evaluated. No obvious difference was observed between the shapes of the average surface reflectance retrieved from the original datacube and those from the decompressed datacubes. Based on the best-estimated input model parameters for atmospheric correction, the mean and standard deviation of the surface reflectance over nine specified zones retrieved from both the original datacube and the four decompressed datacubes were computed. Results show that when hypothesis tests were performed at a level of significance $\alpha = 0.05$, the effects of the compression were significant on the retrieved surface reflectance in some spectral bands, especially for the higher compression ratios. The uncertainty caused by the compression at a compression ratio of 42 : 1 was less than 5%, which is smaller than that caused by the uncertainties in the input model parameters under the assumption that the uncertainties in the input model parameters are independent of each other.

In this paper, we will quantify how lossy VQ hyperspectral data compression affects retrieval of red-edge indices for typical end-user applications. The evaluation will first be applied to CASI datacubes and then validated using an airborne visible/infrared imaging spectrometer (AVIRIS) datacube. Section II will describe data sets to be tested. In Section III, two red-edge indices will be described. Section IV will describe the evaluation method. The experimental results and analysis will be presented in Section V. Finally, conclusions will be drawn in Section VI.

II. Data Sets

Hyperspectral data sets used in this paper were acquired by CASI [11] in July 1996, and by AVIRIS [12] in August 1996, from the southern study area located north of Prince Albert, SK, Canada, through the BOREAS project [10]. These have been converted to units of scene reflectance using the imaging spectrometer data analysis system (ISDAS) atmospheric correction software [13]. Three CASI datacubes of 400 pixels $\times$ 800 lines $\times$ 72 bands $\times$ 16 bits were extracted from the BOREAS data set acquired using the CASI sensor, and named after the BOREAS tower sites, since the datacubes are all from small areas near the tower sites. They are called *yjps*, *ojps*, and *fens* for Young Jack Pine Site, Old Jack Pine Site, and Fen Site, respectively.

An AVIRIS datacube of size 614 pixels $\times$ 512 lines $\times$ 205 bands $\times$ 16 bits was extracted from two data sets that were collected over approximately the same ground targets as the CASI data sets. It is composed of the bottom half of one AVIRIS scene and the top half of the next scene. The datacube is combined in this way so as to include both the Young Jack Pine and Old Jack Pine Sites, which appear in the CASI datacubes. The original 224 AVIRIS bands were reduced to 205 bands by removing bands which are affected by high atmospheric absorption, as the atmospheric correction cannot address the low signal level in these bands.

III. Red-Edge Indices

Two red-edge indices were used for evaluation. These are described below.

A. Vogelmann Red-Edge Reflectance Ratio Index (Vog1) [14]:

This index is correlated with pigment, and is given by the expression

$$\mathrm{Vog1} = \frac{R\,(740\ \mathrm{nm})}{R\,(720\ \mathrm{nm})} \tag{1}$$

where $R(\lambda)$ is the surface reflectance at wavelength λ.

B. Inverse Gaussian Red-Edge Spectral Parameters[15]:

The red-edge parameters or the Inverted-Gaussian Model of the reflectance curve between 670 and 780 nm are defined by

$$R(\lambda) = R_s - (R_s - R_o)\exp\left[-\frac{(\lambda - \lambda_o)^2}{2\sigma^2}\right] \tag{2}$$

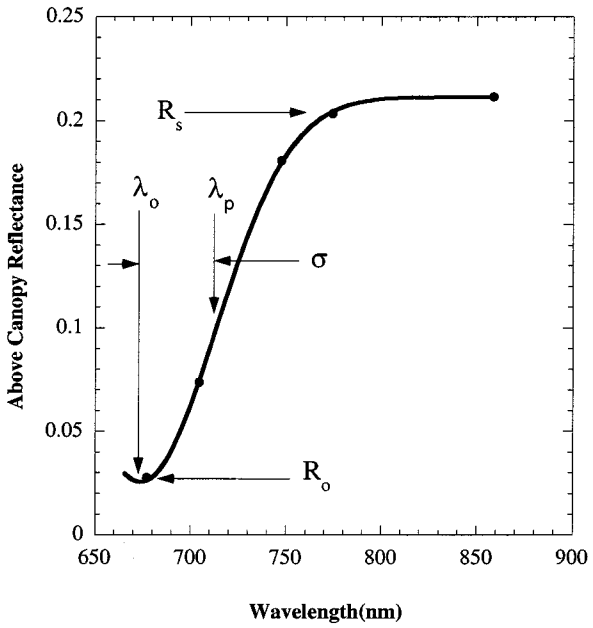

Fig. 1. Definition and curve fit of the inverse Gaussian red-edge parameters.

where

R_s	reflectance maximum;
R_o	reflectance minimum;
λ_o	spectral position of the reflectance minimum;
$\lambda_{\rm p}$	spectral position of the inflection of the Gaussian red-edge reflectance curve;
$\sigma = \lambda_{\rm p} - \lambda_o$	Gaussian curve width parameter.

Fig. 1 depicts the definition of the curve fit and the parameters of vegetation reflectance red-edge in terms of two reflectance parameters (R_s and R_o) and three spectral parameters ($\lambda_{\rm p}$, λ_o and σ).

Both of these two indices [16]–[18] are useful in vegetation applications. They are selected for evaluation because 1) the test datacubes are from vegetated areas, 2) research groups utilizing hyperspectral imagery for forestry applications are currently using the Vog1 index, and 3) the inverse Gaussian red-edge spectral parameters require hyperspectral imagery and cannot be determined from broadband imagery.

In this paper, the following five independent parameters will be computed and examined for evaluation. They are called hyperspectral products in this paper.

1) Vog1: Vogelmann red-edge reflectance ratio index;
2) Red_rs: the reflectance value at the maximum point in the NIR;
3) Red_rd: the difference between reflectance at maximum point in the NIR and reflectance at the minimum point in the red;
4) Red_lo: the wavelength at which the reflectance minimum occurs;
5) Red_lp: the wavelength of the inflection point of the Gaussian fit.

IV. Evaluation Method

This section describes the hyperspectral data compression systems to be evaluated, gives the statistical measures used to show the accuracy of the products due to compression, and finally, considerations for proper retrieval of the products.

A. Data Compression Systems

The 3DVQ algorithm is the basic compression system in our VQ-based hyperspectral data compression series. It does not use any technique for improving the speed and yields the best PSNR. It was used as a reference in the development of improved systems in order to compare performances. In this paper we use this reference system to examine how lossy VQ hyperspectral data compression impacts the selected remote sensing applications. In [7], three improved lossy VQ hyperspectral data compression systems, which integrate the MSCA [6], training set subsampling [5] and the SFBBC [4], have been simulated and tested. *System 1* uses the MSCA with 2% training set subsampling. System 2 uses the MSCA with 2% training set subsampling and SFBBC for codebook training. System 3 uses the MSCA with 2% training set subsampling and SFBBC for both codebook training and for coding. The experimental results show that the overall processing speed of the systems can be improved by a factor of around 1000 at an average PSNR penalty of 1.5 dB. They are the best configuration systems of such lossy VQ hyperspectral data compression techniques based on the MSE distortion measure. In this paper, they are selected for evaluation to examine how the improved systems further impact specific remote sensing applications. Table I shows the configuration of these compression systems.

B. Product Statistics Measures

In this paper, the standard deviation of percentage difference between product values retrieved from the original and from the decompressed data (see (3), shown at the bottom of the page) is used to evaluate the product error induced by lossy compression. Where $\hat{x}_i$ is the product value at pixel i retrieved from the decompressed data, and x_i is the product value at pixel i retrieved from the original data, N is the number of pixels in the scene of a datacube.

C. Computation of Red-Edge Products

The product algorithms are applied to the original and decompressed datacubes on a pixel-by-pixel basis. Since the red-edge

$$\text{Std\_Dev} = \sqrt{\frac{1}{N-1}\left\{\sum_{i=1}^{N}\left(100\frac{\hat{x}_i - x_i}{x_i}\right)^2 - \frac{1}{N}\left(\sum_{i=1}^{N}100\frac{\hat{x}_i - x_i}{x_i}\right)^2\right\}} \tag{3}$$

TABLE I
CONFIGURATION OF THE REFERENCE AND THE THREE IMPROVED HYPERSPECTRAL DATA COMPRESSION SYSTEMS (MSCA STANDS FOR THE MULTIPLE SUBCODEBOOK ALGORITHM, SFBBC STANDS FOR THE SPECTRAL FEATURE-BASED BINARY CODING)

Technique used	Reference	System1	System 2	System 3
Basic 3DVQ	Y	Y	Y	Y
MSCA	N	Y	Y	Y
Training set sub-sampling (2%)	N	Y	Y	Y
SFBBC codebook training	N	N	Y	Y
SFBBC coding	N	N	N	Y

TABLE II
VIABLE DATA RANGE OF THE FIVE RED-EDGE PRODUCTS

Product Name	Minimum Accepted Value	Maximum Accepted Value
Vog1	0.0	10.0
Red_rs [%]	0.0	60.0
Red_rd [%]	0.0	60.0
Red_lo [nm]	660.0	700.0
Red_lp [nm]	693.0	733.0

indices are only valid for vegetated pixels, the products will be computed in a region dominated by vegetation cover in order to reduce the influence of nonvegetated pixels on the evaluation. In the CASI datacubes, a subscene of 256×256 pixels within the scene of a test datacube was selected for computation of the products. The subscene selected contains over 90% vegetated pixels.

Two additional mechanisms were used to prevent computations from nonvegetated pixels in the subscene from impacting on the results: 1) the product algorithms perform "reality checks" on all pixels in a subscene. If a pixel is nonvegetated, it is flagged and will not be used in the subsequent analysis, and 2) the statistics computation algorithm ignores pixels for which the product value does not lie in an expected range. These ranges are given in the Table II.

During the generation of the statistics, the algorithm also records 1) the number of pixels for which both the original datacube and the decompressed datacube produced viable product values; 2) the number of pixels for which neither the original datacube nor the decompressed datacube produced viable product values; and 3) the number of pixels for which only one of the original datacube or the decompressed data produced viable product values (tracked separately). These pixel counts are also used to ensure that the above screening methods are working properly.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the evaluation is first applied to three CASI hyperspectral datacubes and then validated using an AVIRIS datacube.

A. CASI Datacubes

The three CASI test datacubes were compressed using the four systems described in Section IV at eight different codebook sizes. Each system yields the same compression ratio (CR) but a different reconstruction fidelity for the same size codebook. Table III lists the compression ratios corresponding to each codebook size. The compression results are shown in Fig. 2. The definition of PSNR used in this paper is given in [4]. The reference system yields the best PSNR for each of the test datacubes, but it is the slowest. Systems 1 and 2 produce similar PSNRs. Their PSNRs are close to those of

TABLE III
EIGHT CODEBOOK SIZES AND THEIR CORRESPONDING COMPRESSION RATIOS (CR) ON THE CASI TEST DATACUBES

Codebook Size	32	64	128	256	512	1024	2048	4096
CR	225.2	184.9	154.4	129.1	106.2	84.2	62.7	43.1

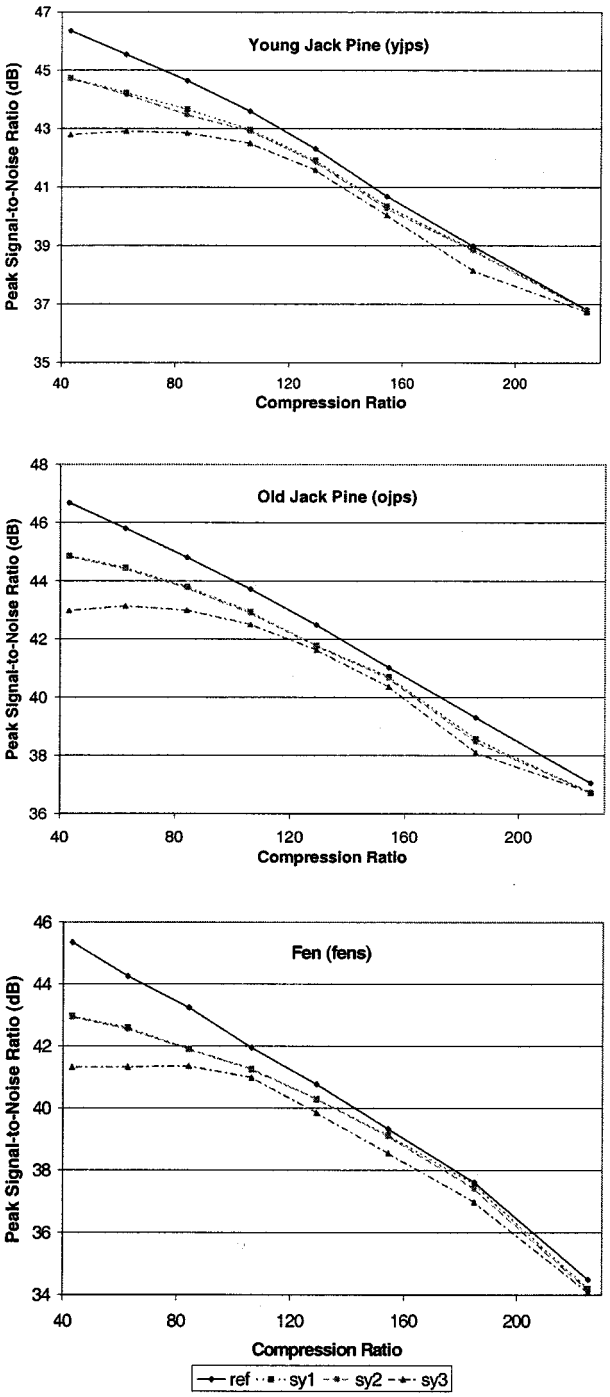

Fig. 2. Compression ratio versus PSNR of the four compression systems on three CASI datacubes.

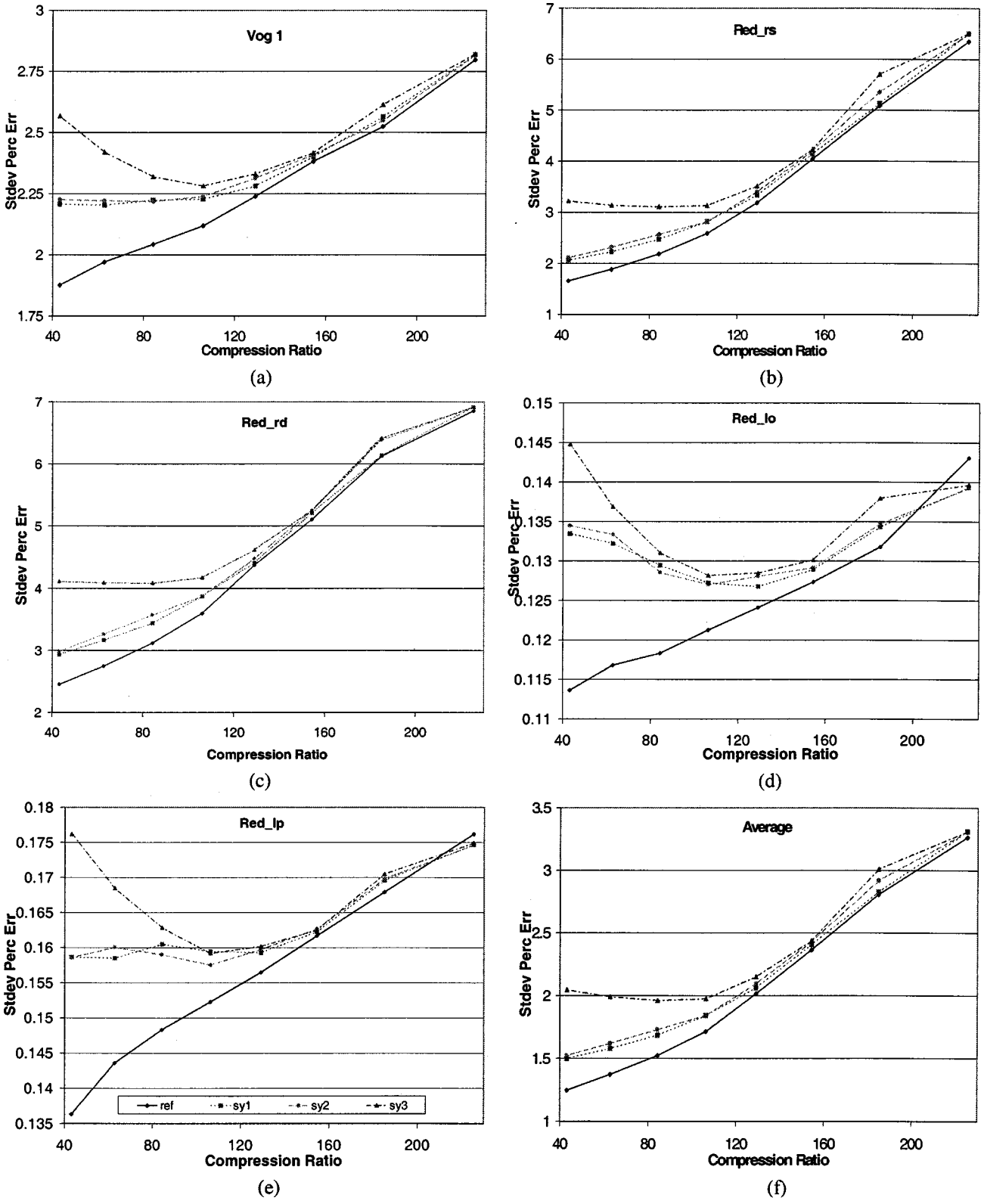

Fig. 3. Standard deviation of percentage difference between products retrieved from the original and decompressed datacubes. The test datacube: Young Jack Pine.

the reference system. System 3 produces the worst PSNR of the systems, but it is the fastest. The compressed data were also decompressed to produce the reconstructed datacube for evaluation.

Five red-edge products: Vog 1, Red_rs, Red_rd, Red_lo and Red_lp were retrieved from each original CASI datacube and from its decompressed datacubes, which were produced by the four compression systems at eight compression ratios. Thus, a

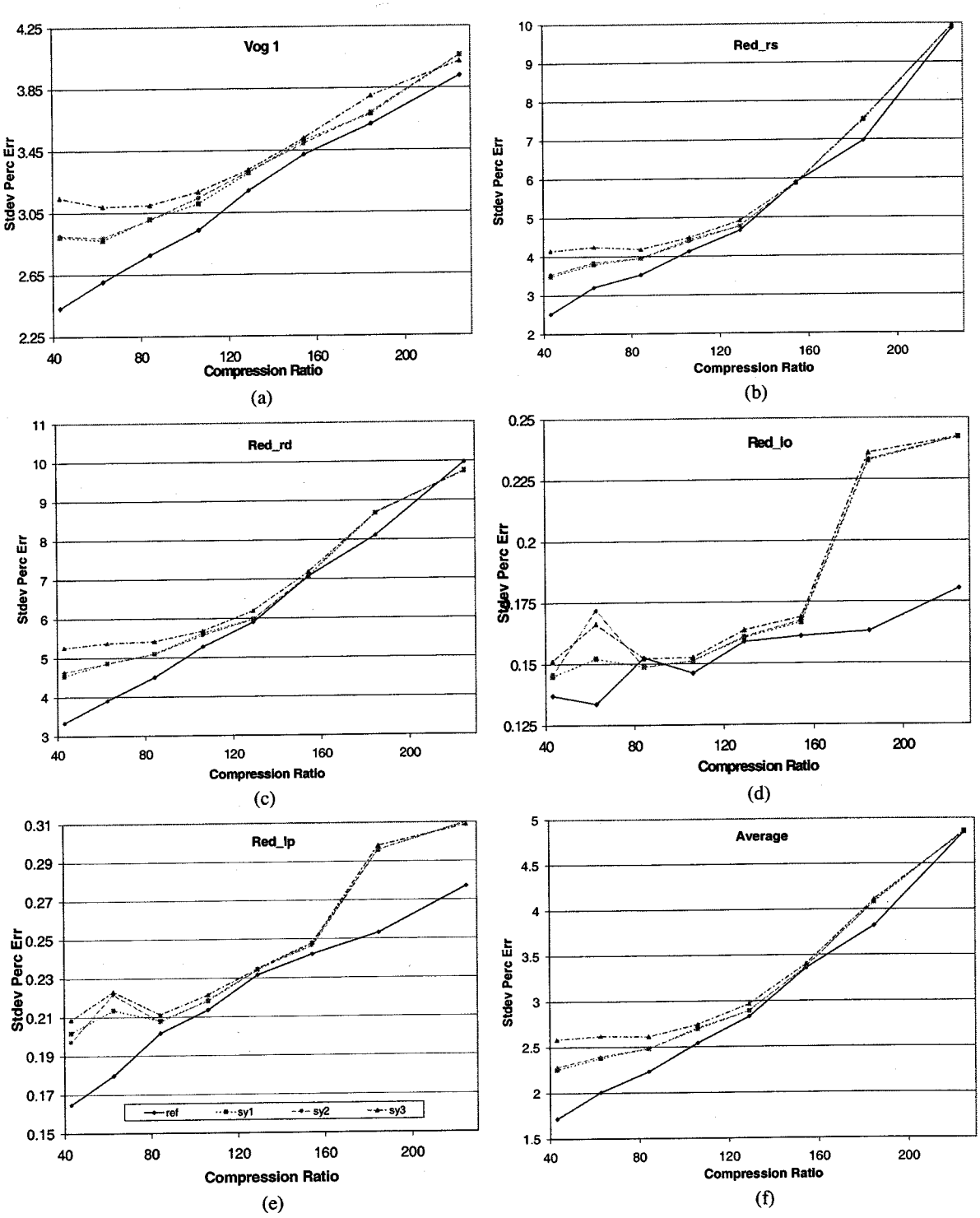

Fig. 4. Standard deviation of percentage difference between products retrieved from the original and decompressed datacubes. The test datacube: Fen.

total of 33 product images were generated per product, since there are 32 decompressed datacubes per original datacube. The Standard deviation of percentage difference between the product retrieved from the original and those retrieved from the decompressed data were computed based on (3).

Fig. 3 shows graphs of standard deviations of percentage difference of the five red-edge products retrieved from the Young Jack Pine Site datacube and its decompressed datacubes and their average graph, as a function of compression ratios. Each curve in a graph represents one compression system, and has eight points, each of which corresponds to a compression ratio. In general, the reference system induces the smallest product errors of the four compression systems. System 1 and 2 perform similarly. Their curves are closer to the reference than to system 3. The product errors increase with increase in compression ratio.

The dynamic range of the standard deviation for Vog 1 is from 1.8% to 2.8%. The dynamic range of the standard deviation of the two reflectance related products is from 1.7% to 7.0%. The dynamic range of the standard deviation of the two wavelength related products is very small. It is smaller than 0.14%. This indicates that the lossy VQ compression has little impact on wavelength. The spectral information in the wavelength domain is well preserved during the process of this lossy compression. The curves of the product errors corresponding to different compression systems start to converge once the compression ratio exceeds 106.

In the Vog 1 graph, the standard deviations from compression systems 1 and 2 remain constant (around 2.2%) for compression ratios from 43 to 106. In contrast, the standard deviation from the reference system increases from 1.8% to 2.1% at the same compression ratios. In this case, system 1 and 2 with a compression ratio of 106 are optimal from the point of view of the Vog 1 application, since they produce the best compression and induce the smallest error on the product. The curves of standard deviation versus compression ratio are almost linear once the compression ratio exceeds 106. The standard deviation values from compression system 3 at compression ratios 43, 62, and 84 are worse than that at compression ratio 106. This is probably caused by the instability of the compression system. It can be seen from the PSNR versus CR curves shown in Fig. 2 that the improvement of PSNR becomes very small when the compression ratio decreases below the compression ratio of 106. For a compression ratio of 43, the PSNR of system 3 is actually worse than it is at slightly higher compression ratios. Comparing the Vog 1 curve with the PSNR curve, we see that they demonstrate similar trends. The only difference is in their sensitivity. Vog 1 is more sensitive than PSNR to this instability of the system. Fortunately, the absolute amount resulting from this instability is small. In this example it is less than 0.3%.

In the Red_rs and Red_rd graphs, the standard deviations of the systems are relatively large when the compression ratio is high. These almost reach 6.5% and 7.0% in graphs Red_rs and Red_rd, respectively, when the compression ratio is 225. They are more sensitive to compression than other products. The curves of standard deviation from compression systems 1 and 2 are very close to that from the reference system in both of the graphs. The product errors from system 3 remain constant at compression ratios from 43 to 106 in both graphs. They are slightly above 3.0% and 4.0% in the two graphs. Below a compression ratio of 106, this is caused by the poorer performance of system 3 compared to the other systems.

In the two wavelength-related products, graphs Red_lo and Red_lp, all the three speed-improved compression systems perform poorly with respect to the reference system when the compression ratio is smaller than 106. In the graph Red_lo, the standard deviation curves from these three systems bend upward when the compression ratio is less than 106. The product errors decrease with the increase of compression ratio up to 106 rather than increase. In the graph Red_lp, the standard deviations from system 3 are similar to those in graph Red_lo, while the standard deviations from systems 1 and 2 slightly oscillate around a value of 0.16%. These results indicate that at a lower compression ratio (< 106), the wavelength related products are more sensitive to the improved compression techniques than the reflectance related products, although the absolute value of the errors is 15 to 40 times smaller than that of the reflectance products. System 3 performs poorest in the three improved systems with respect to the reference system.

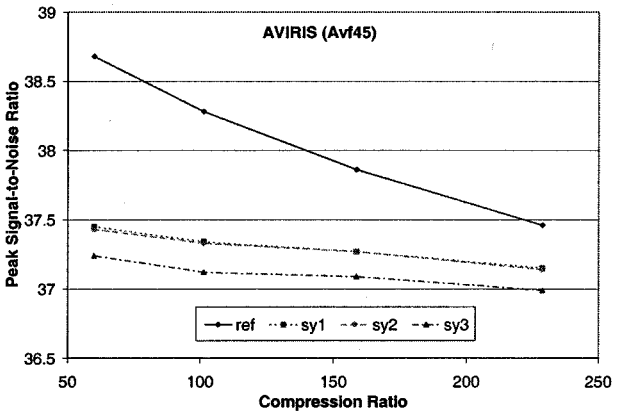

Fig. 5. Compression ratio versus PSNR of the four compression systems on AVIRIS datacube.

Graph (f) in Fig. 3 is the average over the five products. It is presented to summarize the overall product error of a test datacube. The average product errors are dominated by products Vog 1, Red_rs, and Red_lp, which have large values of standard deviation. It can been seen from this graph that system 1 and 2 perform as well as the reference system in terms of the overall product errors. System 1 and 2 compress a datacube hundreds of times faster than the reference system. The differences between the overall errors from the reference and those from system 1 or 2 are below 0.25% at all compression ratios. They are insignificant, especially when compression ratios are greater than 106. It is recommended that system 1 or 2 be used with a compression ratio of no more than 106, if one wants to obtain a relatively small error on these products. In this way, the overall product error would be below 2.0%. This is an acceptable error level in most of applications.

For example, spectral indices used for the estimation of chlorophyll content in higher plant leaves are typically based on the red-edge inflection point Red_lp or reflectance ratios such as R_{750}/R_{700} and R_{750}/R_{550} [19], [20]. A 2.0% error in the retrieved reflectance ratio could produce an error in the reflectance ratio of 4.0%. The resulting uncertainty of the estimated chlorophyll content Chl $a + b$ using the results of [19] for maple leaves for example, would be 0.56 μg cm^{-2}. This is much less than the estimated error in the regression using the R_{750}/R_{700} ratio of 3.6 μg cm^{-2}.

Each of the above experiments was also carried out on the Old Jack Pine datacube. The results showed similar trends, except the results are slightly better than those of the Young Jack Pine datacube, since the PSNRs obtained using the Old Jack Pine datacube are slightly better than those obtained using the Young Jack Pine datacube. For brevity, they are not presented here.

Fig. 4 shows the graphs of standard deviations of the products retrieved from the Fen datacube. The trends of standard deviations of the products Vog 1, Red_rs, and Red_rd are similar to those of the Young Jack Pine datacube, but their values

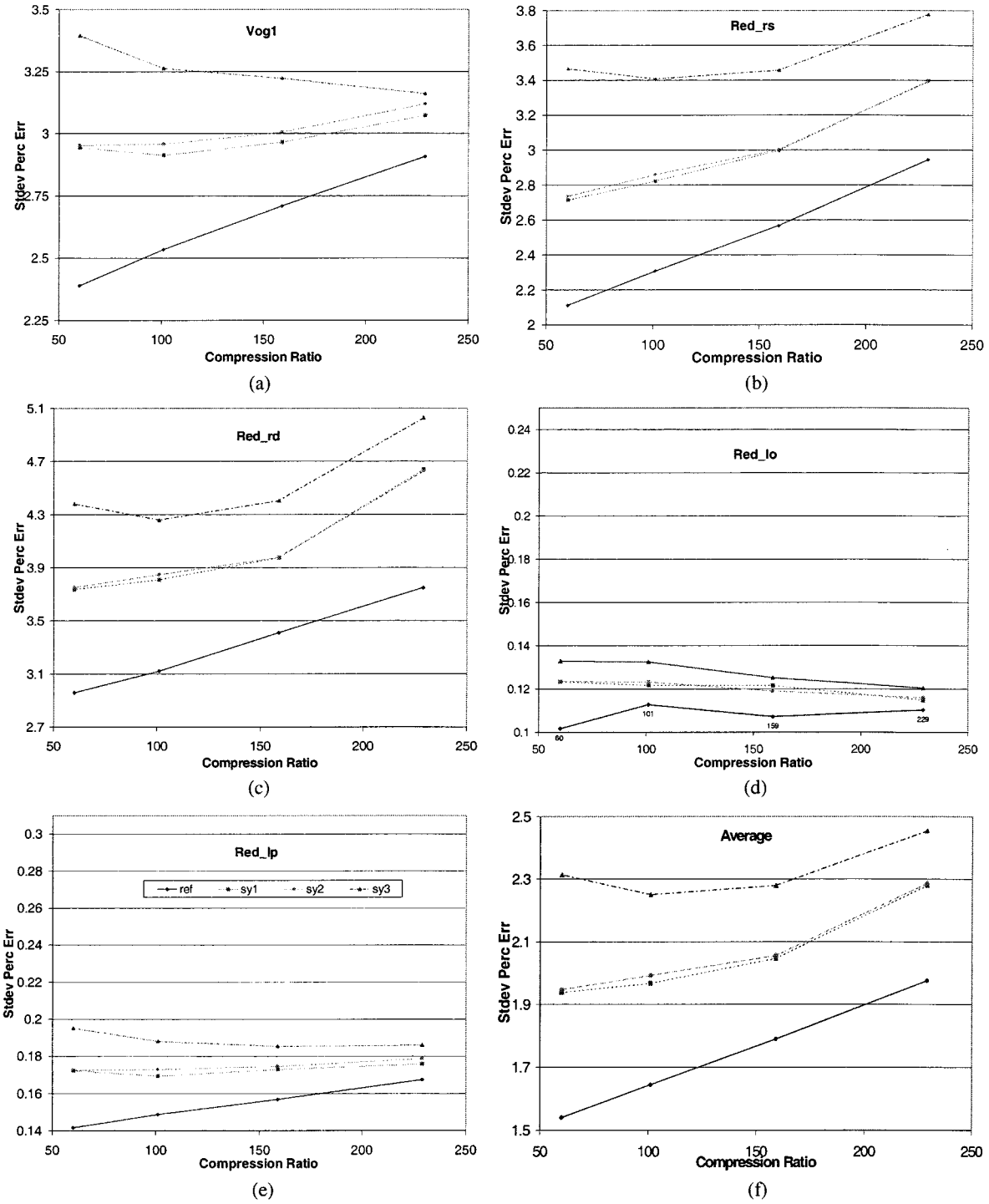

Fig. 6. Standard deviation of percentage difference between products retrieved from the original and from decompressed datacube. The test datacube: AVIRIS.

are slightly worse than those for the Young Jack Pine datacube (1.0% to 3.0% worse). This is because the PSNRs of the Fen decompressed datacubes are about 1.5 dB lower than those of Young Jack Pine datacube (see Fig. 2).

In general, products Red_lo and Red_lp from the Fen Site are up to 0.14% worse than those from the Young Jack Pine Site. As mentioned previously, this is caused by slightly poorer PSNR of the decompressed datacubes. The Red_lo and Red_lp curves of the reference system from the Fen Site are not as straight as those from the Young Jack Pine Site, especially the Red_lo curve. Unlike the Young Jack Pine Site, the Red_lo and Red_lp curves from the Fen Site compressed by the three improved systems are very close and do not bend upward explicitly in the low compression ratio range. The differences between standard deviations from the improved systems and those from the reference are around 0.07% when compression ratios are 184 and 225 in the two products. They are relatively large compared to the difference of 0.006% at compression

ratios between 84 and 154, although the absolute value is quite small.

The average standard deviations over the five products are plotted in graph (f) of Fig. 4. Similar to the Young Jack Pine Site, they are dominated by products Vog 1, Red_rs and Red_lp. System 1 and 2 perform as well as the reference system. The differences between overall product errors from the reference system and from system 1 or 2 are below 0.5% at all compression ratios. For the Fen datacube, system 3 performs quite well. Compared with the Young Jack Pine datacube, the degradation of performance in the low compression ratio range is not explicit.

B. AVIRIS Datacube

The AVIRIS datacube was compressed using the four systems at four codebook sizes: 512, 1024, 2048, and 4096. With these four codebook sizes, the four systems yield the same compression ratios of 229, 159, 101, and 60, respectively. These compression ratios are selected for evaluation since they are in a similar range to those obtained using the CASI datacubes. Fig. 5 shows the compression results in terms of the MSE distortion measure. The PSNRs of the decompressed data, produced by these four systems on the AVIRIS datacube, are smaller than those on the CASI datacubes. The decrease of PSNR with the increase of compression ratio is also small. The PSNR produced by the reference system decreases only 1.3 dB from the smallest compression ratio of 60 to the largest ratio of 229. The PSNRs produced by the improved systems are very close. System 1 and 2 produced almost the same PSNR at the four compression ratios, so that their compression ratio versus PSNR curves are merged together. The maximum difference between PSNRs produced by system 1 or 2 and by system 3 is only 0.2 dB.

The product errors retrieved from the original AVIRIS datacube and its decompressed datacubes, as a function of compression ratio, is shown in Fig. 6. Compared with the graphs in Figs. 3 and 4, the graphs of the product errors from the AVIRIS datacube demonstrate similar trends to those obtained from the CASI datacubes. In general, the standard deviations of all the products from the AVIRIS datacube are smaller than those from the CASI datacubes. This probably results from the attributes of the AVIRIS datacube. Most notably, the AVIRIS datacube is approximately three times larger than the CASI datacubes. Therefore, higher compression ratios are expected. For compression ratios near 106, the CASI Fen and Young Jack Pine datacubes yield error statistics similar to the error statistics from the AVIRIS datacube. For compression ratios above 106, the overall product error increase for the AVIRIS datacube is minimal, while the overall product error increase for the CASI datacubes is quite significant. To achieve comparable compression ratios, the codebook size for the CASI datacubes must be much smaller than for the AVIRIS datacube (for example, to achieve a compression ratio 106, the CASI codebooks can only contain 512 codevectors, while the AVIRIS codebooks can contain 2048 codevectors). Therefore, the impact of compression ratio on overall product error is more significant for the CASI datacubes than it is for the AVIRIS datacube. Fig. 7 shows the average product errors of the CASI and AVIRIS datacubes compressed by system 2 for the scenario detailed above.

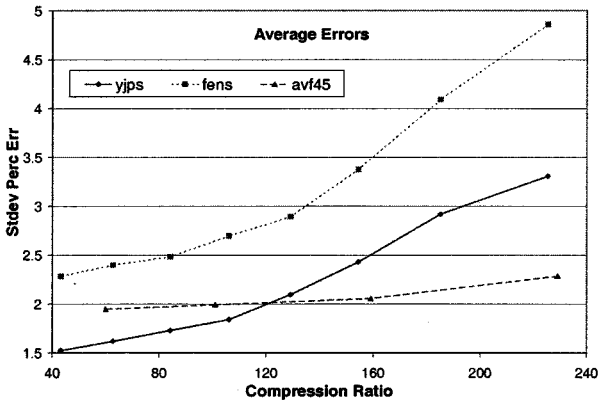

Fig. 7. Comparison of overall product errors from CASI datacubes (Young Jack Pine and Fen) and AVIRIS datacube compressed by system 2.

The CASI datacubes selected for evaluation in this paper demonstrate worst case results because the datacubes are small. If a datacube is large enough spatially, a high compression ratio can be obtained even if a large codebook is used, and this will yield minimal induced product errors. This validation test gives us more confidence on the product error statistics.

C. Spatial Patterns in Induced Product Errors

The previous analysis shows statistical results from comparisons between data products derived from original and decompressed datacubes. In this section, the spatial component of the product errors is presented. Fig. 8 shows a series of pictures of spatial patterns of the product errors retrieved from the AVIRIS datacube. A quick-look image of the complete scene of the AVIRIS datacube, generated at 733 nm, is presented in picture Fig. 8(a). Pictures Fig. 8(b)–(f) show the absolute value of the percentage error between each red-edge product calculated from the original datacube and from the decompressed datacube produced by system 3. System 3 yields the worst PSNR in terms of MSE distortion and largest product errors of the four compression systems. Thus the pictures of spatial patterns resulting from system 3 show the worst cases of the compression systems.

These pictures show that for all products, errors are relatively high along the north-south river and river bank and along narrow logging roads, where mixed pixels would occur. For all products, errors are relatively high in clear-cut areas. This would be in part because the area is primarily nonvegetated (some limited regrowth only) and because the regions tend to be nonuniform and would therefore contain highly mixed pixels consisting of various amounts of soil, deadfall, and a variety of vegetation in different stages of growth. For the Vog1 product, the highest errors are generated over water bodies (the bright regions in picture Fig. 8(b). The high errors for nonvegetated pure and mixed pixels are inherently due to the fact that the red-edge algorithm does not apply to the spectra for these pixels. Consequently this error is of no significance. In the area of interest (vegetation) the error is uniformly distributed.

VI. Conclusion

In this paper, four lossy VQ hyperspectral data compression algorithms were evaluated using red-edge indices as end-prod-

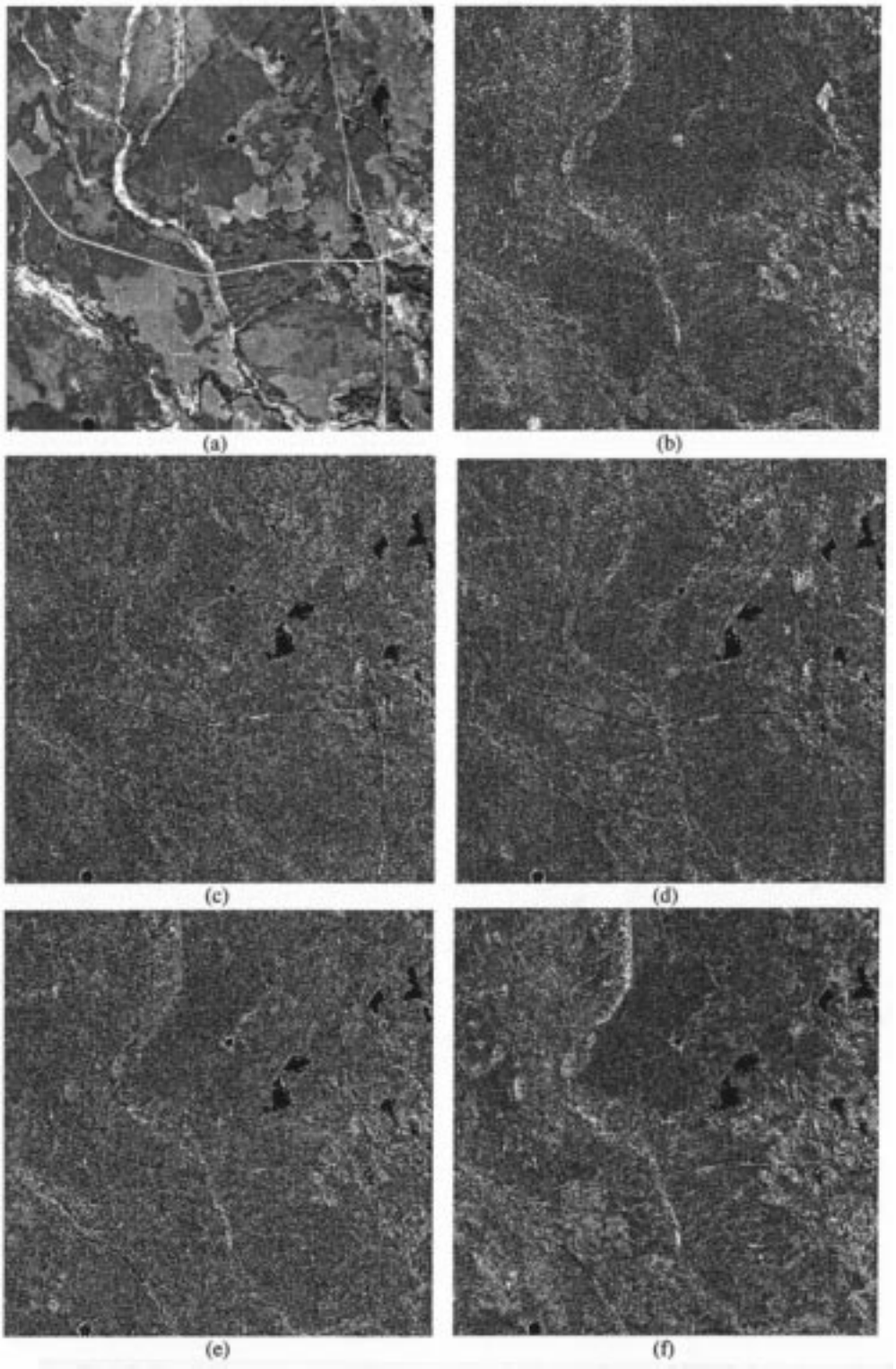

Fig. 8. Spatial patterns in induced product errors of the AVIRIS data. (a) Quick-look image of the scene at wavelength 733 nm. (b)–(f) Images of the absolute value of the percentage error between each red-edge product calculated from the original datacube and from the reconstructed datacube compressed by system 3 in order of products: Vog1, Red_rs, Red_rd, Red_lo, and Red_lp.

ucts. The standard deviation of percentage difference between a product retrieved from an original datacube and that from its decompressed datacube was used as a measure to quantify the impact of compression on end-products. Three CASI datacubes from vegetated areas were tested. An AVIRIS datacube collected over approximately the same ground targets at the approximately the same time of the year was also tested to validate the experimental results.

Four lossy VQ hyperspectral data compression systems were examined. A basic compression system called the reference system, which does not use any technique for improving the speed, yielded the best PSNR for all test datacubes. But it is

the slowest. Systems 1 and 2 produced PSNR close to that of the reference. System 3 produced the worst PSNR, but it is the fastest. Five red-edge products: Vog 1, Red_rs, Red_rd, Red_lo, and Red_lp were retrieved from each original datacube and from their decompressed datacubes, and were analyzed.

For CASI datacubes, the reference induces the smallest product errors of the four compression systems. Systems 1 and 2 perform fairly similarly to the reference. System 3 performs similarly to the reference at high compression ratios. Product errors increase with the increase of compression ratio. The amplitude of product error of the two wavelength-related products Red_lo and Red_lp (below 0.18%) is 15 to 40 times smaller than that of the two reflectance related products Red_rs and Red_rd. This indicates that the lossy VQ compression has little impact on wavelength. The spectral information in wavelength domain can be well preserved during the process of lossy compression. The overall product errors are dominated by Vog 1, Red_rs and Red_lp, which have relatively large value of induced error. The difference between the overall error from the reference and that from system 1 or 2 is below 0.5% at all compression ratios. This is insignificant. Systems 1 and 2 compress a datacube hundreds of times faster than the reference system. The overall product error induced by systems 1 or 2 is below 2.0% for the Young Jack Pine and Old Jack Pine datacubes and below 3.0% for the Fen datacube when the compression ratio is 106 and below. It is recommended that system 1 or 2 be used with a compression ratio not greater than 106 if one wants to obtain a relatively small error on these products.

For the AVIRIS datacube, similar trends as in the CASI datacubes were observed. In general, the product errors from the AVIRIS datacube are smaller than those from the CASI datacubes. This probably results from the greater compressibility of the AVIRIS datacube, which is due to the fact that it is three times larger than the CASI datacubes and thus, higher compression ratios are expected. When the compression ratio reaches a certain value, the CASI codebooks can contain only a small number of codevectors to achieve a compatible compression ratio while the AVIRIS codebooks can contain a large number of codevectors. Minimal induced product errors would be expected if a datacube is large enough spatially, since a codebook can still contain a sufficiently large number of codevectors at a high compression ratio.

Spatial patterns of the product errors of the AVIRIS datacube were also presented. For all products, errors are uniformly distributed in vegetated areas. Errors are relatively high in nonvegetated and mixed-pixel areas.

REFERENCES

[1] R. M. Gray, "Vector quantization," *IEEE Acoust., Speech, and Signal Processing Mag.*, vol. 1, pp. 4–29, 1984.

[2] S.-E. Qian, A. B. Hollinger, D. Williams, and D. Manak, "3D data compression systems based on vector quantization for reducing the data rate of hyperspectral imagery," in *Applications of Photonic Technology 2*. New York: Plenum, 1997, pp. 641–654.

[3] ——, "A near lossless 3-dimensional data compression system for hyperspectral imagery using correlation vector quantization," in *Proc. 47th Int. Astronautical Congr.*, Beijing, China, Oct. 1996.

[4] ——, "Fast 3D data compression of hyperspectral imagery using vector quantization with spectral-feature-based binary coding," *Opt. Eng.*, vol. 35, no. 11, pp. 3242–3249, 1996.

[5] D. Manak, S.-E. Qian, A. B. Hollinger, and D. Williams, "Efficient hyperspectral data compression using vector quantization and scene segmentation," *Can. J. Remote Sensing*, vol. 24, pp. 133–143, June 1998.

[6] S.-E. Qian, A. B. Hollinger, D. Williams, and D. Manak, "3D data compression of hyperspectral imagery using vector quantization with NDVI-based multiple codebooks," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, Seattle, WA, July 1998, pp. 2680–2684.

[7] ——, "Vector quantization using spectral index based multiple sub-codebooks for hyperspectral data compression," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1183–1190, May 2000.

[8] M. J. Ryan and J. F. Arnold, "Lossy compression of hyperspectral data using vector quantization," *Remote Sens. Environ.*, vol. 61, pp. 419–436, 1997.

[9] B. Hu, S.-E. Qian, and A. B. Hollinger, "Impact of lossy data compression using vector quantization on retrieval of surface reflectance from casi imaging spectrometry data," *Can. J. Remote Sensing*, vol. 27, pp. 1–19, Feb. 2001.

[10] J. R. Miller, J. R. Freemantle, P. R. Shepherd, L. Gray, N. T. O'Neill, A. Royer, and E. Senese, "Deployment of casi to meet the needs of BOREAS science," in *Proc. 17th Can. Remote Sensing Symp.*, Saskatoon, SK, 1995, pp. 169–175.

[11] C. D. Anger, S. Mah, T. A. Ivanco, S. B. Achal, R. Price, and J. Busler, "Extended Operational Capabilities of the casi," in *Proc. 2nd Int. Airborne Remote Sensing Conf. Exhibition*, vol. 1, San Francisco, CA, 1996, pp. 124–133.

[12] G. Vane, R. O. Green, T. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 44, pp. 127–143, 1993.

[13] K. Staenz, "ISDAS—A system for processing and analysing high spectral resolution data with emphasis on surface reflectance retrieval,", [Online] Available: http://ltpwww.gsfc.nasa.gov/ISSSR-5/isdasasy.htm.

[14] J. E. Vogelmann, B. N. Rock, and D. M. Moss, "Red-edge spectral measurements from sugar maple leaves," *Int. J. Remote Sensing*, vol. 14, no. 8, pp. 1563–1575, 1993.

[15] G. F. Bonham-Carter, "Numerical procedures and computer program for fitting an inverted Gaussian model to vegetation reflectance data," *Comput. Geosci.*, vol. 14, no. 3, pp. 339–356, 1988.

[16] J. R. Miller, J. Wu, M. G. Boyer, M. Belanger, and E. W. Hare, "Seasonal patterns in leaf reflectance red-edge characteristic," *Int. J. Remote Sensing*, vol. 12, no. 7, pp. 1509–1523, 1991.

[17] J. R. Miller, E. W. Hare, and J. Wu, "Quantitative characterization of the vegetation red edge reflectance 1. An inverted-Gaussian reflectance model," *Int. J. Remote Sensing*, vol. 11, no. 10, pp. 1755–1773, 1990.

[18] M. Belanger, J. R. Miller, and M. G. Boyer, "Comparative relationships between some red edge parameters and seasonal leaf chlorophyll concentrations," *Can. J. Remote Sensing*, vol. 21, no. 1, 1995.

[19] A. A. Gitelson and M. N. Merzlak, "Remote estimation of chlorophyll content in higher plant leaves," *Int. J. Remote Sensing*, vol. 18, no. 12, pp. 2691–2697, 1997.

[20] H. K. Lichtenthaler, A. Gitelson, and M. Lang, "Non-destructive determination of Chlorophyll content of leaves of a green and an aurea mutant of Tobacco by reflectance measurements," *J. Plant Physiol.*, vol. 148, pp. 483–493, 1996.

Shen-En Qian received the B.Eng. degree in industrial electrical automation from Hefei University of Technology, Hefei, China, in 1982, the M.S. degree in opto-electronics from the Chinese Academy of Sciences, Beijing, in 1985, and the Ph.D. degree in communication and electronic systems from Jilin University of Technology, Jilin, China, in 1990.

He is currently a Senior Research Scientist with the Space Technologies Division, Canadian Space Agency, St. Hubert, PQ, where he has been since 1995. Previously, he was a Professor with the Changchun Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, and was with the Departement de Recherche Spatiale, Observatoire de Paris-Meudon, CNRS, Paris, France for one and a half years. He has published over 60 papers in the areas of opto-electronic signal detection and reception, weak-signal detection, intelligent instrumentation, real-time signal processing, image analysis, and processing for remotely sensed data and hyperspectral data, and data compression.

Dr. Qian was a recipient of the 1992 Marie Curie Award (European Community International Scientific Cooperation Program). He received the 1999 Director Award of the Federal Government of Canada.

Allan B. Hollinger received the Ph.D. in physics from University of Toronto, Toronto, ON, Canada, in 1977.

He is the Manager for Sensors and Signal Processing, Space Technologies Division, Canadian Space Agency, St. Hubert, PQ. From 1979 to 1988, he was with Moniteq Ltd., where he was part of the team that first developed the fluorescence line imager imaging spectrometer. This airborne sensor for a variety of remote sensing applications. From 1988 to 1993, he was Manager of the Electro-Optics Laboratory, Institute for Space and Terrestrial Science, where he and his colleagues developed protocols for the precision radiometric calibration of imaging spectrometers. He is currently focusing his attention on the development of space technology and standardized applications of imaging spectrometers to solving remote sensing problems. He has contributed to over 50 technical reports and publications describing the development and application of electro-optical instrumentation for remote sensing.

Melanie Dutkiewicz received the B.Sc. degree in physics and engineering from the University of Cape Town, Cape Town, South Africa, in 1983 and 1985, respectively. She received the Ph.D. degree in engineering from Cambridge University, Cambridge, U.K., in 1990.

She joined MacDonald Dettwiler and Associates (MDA), Richmond, BC, Canada, as a part of the Research and Development Department. During her time with MacDonald Dettwiler, she worked on a large variety of projects, including compression of synthetic aperture radar (SAR) data for both on-ground and on-board systems, calibration and analysis of polarimetric SAR data, glacier flow rate detection from satellite data, context-sensitive remote sensing data fusion, and hyperspectral data compression, processing, and analysis.

Herbert Tsang received the B.S. degree in electrical engineering from Southern Illinois University, Carbondale, in 1993, and the M.S. degree in electrical engineering from the University of Washington, Seattle, in 1995.

He joined MacDonald Dettwiler and Associates (MDA), Richmond, BC, Canada, in 1996 as a Member of the scientific and engineering staff. During his tenure at MDA, he worked in a large variety of space and defense-related projects in the research and development department. His research interests include system simulation and optimization, digital image processing, confocal microscopy imaging, and hyperspectral imaging.

Mr. Tsang is a member of the Engineering in Medicine and Biological Society and is a registered Professional Engineer in the province of British Columbia, Canada.

Harold Zwick received the B.A., M.A., and Ph.D. degrees in physics from the University of Saskatchewan, Saskatoon, SK, Canada.

He joined MacDonald Dettwiler and Associates (MDA), Richmond, BC, Canada, in 1986 as a Member of the Scientific and Engineering Staff. At MDA, he has led a number of research and development projects in Earth observation. His main interest has been in sensors, sensor systems, and remotely derived information products.

James R. Freemantle received the B.Sc. degree in physics from the University of Waterloo, Waterloo, ON, Canada, and the M.Sc. degree in physics from York University, Toronto, ON, Canada.

Since 1990, he has been a Project Scientist with the Earth Observations Laboratory, Centre for Research in Earth and Space Technology (CRESTech), York University. His research interests include atmospheric correction of airborne and satellite imagery, software development for the processing of airborne imagery from compact airborne spectrographic imagers, and real-time analysis of remote sensing data.